提高开放发现透明度：美国 NISO ODI 述评*

胡小丽[1] 刘铁威[2] 熊丽[1] 薛燕[3]
[1] 首都师范大学图书馆 北京 100089 [2]加州州立大学弗雷斯诺分校图书馆 弗来斯诺 93740－8014
[3] 加州大学伯克利分校东亚图书馆 伯克利 94720－6000

摘要：[**目的/意义**]对美国国家信息标准协会(NISO)发布的"开放发现倡议:提高发现的透明度"(ODI)进行研究,借此推动我国图书馆界对ODI项目的深入了解与广泛应用,促进基于中心索引仓储的发现服务的发展和数字资源建设的标准化发展。[**方法/过程**]介绍ODI产生的背景、发展状况、核心内容与特点,同时对该规范的最佳实践指南和相关组织的活动进行总结。在此基础上调研国内外发现服务领域对ODI的应用与实践现状,分析ODI的发展趋势。[**结果/结论**]ODI在元数据共享、公平链接、规范数据格式、促进内容交接的一致性、使用统计等方面提出了一系列推荐实施规范,目前已有较多的资源提供商和发现服务提供商认可这些规范并付诸实践,这将有益于促进数字资源的标准化建设。

关键词：ODI 发现服务 中心索引 元数据
分类号：G250.7
DOI:10.13266/j.issn.0252－3116.2018.06.007

chinaXiv:202308.00364v1

开放发现倡议(Open Discovery Initiative,简称ODI)作为基于中心索引仓储的新一代图书馆发现服务的推荐实施规范,通过规范内容资源元数据格式、数据交换格式与传输标准,倡导公平链接,促进元数据共享与内容交换的一致性,提高发现服务透明度,提升图书馆文献发现与传递服务能力,保障最终用户对文献资源的发现与检索效益。

1 NISO ODI 产生的背景

随着数字资源与网络资源逐渐成为馆藏与读者使用的主体资源,如何整合各类资源,满足读者一站式、方便、快捷地了解、查找、发现、使用资源的需求,成为图书馆文献发现与传递服务应用性研究的重要课题。近年,信息化技术的发展,将基于中心索引仓储搜索模式的发现系统引入到图书馆的服务中。

基于中心索引仓储的发现服务是指基于中心索引集合,实现对图书馆所有资源(包括所有来自于不同提供者的订购资源和免费资源)的全覆盖搜索。其索引集合由元数据、全文数据或者其他资源内容的表述数据组成,包括期刊论文、电子书、研究报告、参考资源、图像、地图、数据集、音视频等各种资源类型。全球第一个网络级资源发现系统Summon由ProQuest公司于2009年推出,随后,OCLC推出了WorldCat Local(简称WCL),艾利贝斯(Ex Libris)公司推出Primo Central(简称Primo),EBSCOhost公司推出EBSCOhost Discovery System(简称EDS)。作为新一代的文献发现与传递服务模式,这几个主要发现系统一被投放到市场,立即引起全球图书馆的关注并迅速成为文献发现与传递的主体服务模式[1]。

基于中心索引仓储的发现服务是一个错综复杂的生态系统,在其为图书馆及最终用户在资源检索与发现上带来巨大益处的同时,也产生了一些新问题,对行业实践提出了新挑战。发现服务的推行涉及到内容提供商(content providers)、发现服务提供商(discovery service providers)和图书馆等多个行业机构的利益,关联到元数据规范、数据交换格式、数据传输标准、公平链接、使用统计等一系列相互纠缠、焦灼的问题。

内容提供商包括商业机构和非营利组织,是内容

* 本文系2016年度国家社会科学基金项目"中美电子资源国家标准比较研究"(项目编号:16BTQ027)研究成果之一。
作者简介:胡小丽(ORCID:0000－0003－2724－8228),信息咨询部、系统部主任,副研究馆员,硕士,E-mail:huxiaoli@cnu.edu.cn;刘铁威(ORCID:0000－0003－2455－6487),馆员,硕士;熊丽(ORCID:0000－0002－3377－7487),副馆长,研究馆员,硕士;薛燕(ORCID:00000－0002－0474－7828),馆员,硕士。
收稿日期:2017－09－05 修回日期:2017－11－17 本文起止页码:51－60 本文责任编辑:刘远颖

ChinaXiv合作期刊

产品或服务的供应者,也是发现服务中心索引集构建的主要提供者。发现服务提供商是基于中心索引仓储的发现系统的构建者,图书馆从内容提供商处获得资源,发现服务提供商订购系统平台,进而向其最终用户提供便捷的文献发现与传递服务。

在发现服务运行中,需要依靠内容提供商、发现服务提供商合作提供信息资源的元数据或全文以建立高效实用的索引。而图书馆在整个服务中需要清楚地知道自己所拥有的内容资源在发现服务中的被揭示程度,这是图书馆评估、选择发现系统和持续推进该项服务所必须的。例如,图书馆需要清楚地了解:哪些文章、数字资源被呈现;这些内容是被全文索引,仅提供引文,还是两者皆有;元数据是来自于集成数据库还是文摘索引资源等。

但实际情况是发现服务索引体系中的数据多是基于内容提供者与发现服务提供商之间的私下协议和特别交换模式组建的,其元数据质量、数据交换格式与方法、检索机制、检索结果相关度排序等有失规范或者缺少必要的公开说明。这种情况为三方利益相关者都带来了隐忧,阻碍了发现服务推广的进程。对于图书馆来说,这一情况下,无法清楚地了解发现服务对馆藏资源的覆盖率与揭示程度,无法了解读者真实的文献使用需求与资源使用情况。对于内容提供者而言,虽然基于中心索引仓储的文献发现模式有利于提高其资源的被发现机率和使用率,但是也带来了一些困扰。例如,图书馆用户可以绕开内容提供者提供的专用检索界面,这会潜在地降低或消除其品牌的知名度,并且使内容提供商无法控制内容呈现给最终用户的程度与方式。此外,内容提供商对于发现平台是否可以公平地发现其所提供的内容资源也存有疑虑。对于发现服务提供商而言,一方面苦于索引仓储中元数据的良莠不齐,另一方面烦恼于其链接机制的公平性常常被质疑。发现系统中,众多不同因素决定了内容的揭示方式、检索结果的排序以及用户访问内容资源的方式。但是如果发现服务提供商与内容提供商隶属于同一母公司,或者被发现其与某内容提供商之间存有独家授权的协议或其他业务关系,其检索机制与链接的公平性就会遭到质疑,从而影响其发现平台的被认可度。

上述问题从一开始就伴随着发现服务的诞生而存在,如果没有采取必要的举措进行干预,势必会造成这一服务的停滞。图书馆、内容资源提供者、发现服务提供商等相关行业就此提出提升发现服务透明度和加强标准化建设的要求,NISO ODI 项目应运而生。目前,国内还没有对 ODI 项目的深入引介与研究,希望通过本文的调查研究,能够帮助国内图书馆界更好地理解 ODI 项目,解决发现服务中的实际问题,推动国内外发现服务研究的交流与发展,为数字资源的标准化与规范化建设提供有益参考。

2 NISO ODI 项目发展概况

ODI 项目的始建得益于艾利贝斯(Ex Libris)公司的奥伦·拜特·阿里(O. Beit-Arie)及独立顾问詹妮·沃尔克(J. Walker)和图情界资深顾问马歇尔·布里丁(M. Breeding)等人的努力。在2011年6月美国图书馆协会(American Library Association,简称ALA)新奥尔良年会期间,他们邀请多位业内资深经理人会面,探讨大家共同关切的图书馆发现服务焦点问题。并在首发会议后,正式向 NISO 递交了提案。同年末,NISO 发现到传递专题委员会接受了该项提案并于2012年初成立 ODI 工作组,正式确立了 ODI 项目。ODI 工作组成员由来自于图书馆、内容提供商和发现提供商的代表共同组成,马歇尔·布里丁和詹妮·沃尔克担任联合主席。

为了更好地了解相关行业对新一代发现服务的满意度及评估相关实践需求的紧迫性,制定切实有效的推荐实施规范,NISO 工作组于2012年9月11日–10月4日,在图书馆员、内容提供商和发现服务提供商中进行了跨行业的问卷调研。调研涉及了学术型元数据的传递/索引层次、数据交换技术、改善发现工具使用数据等诸多方面。此次调研联系了相关领域2 000多人,最终收集到有效问卷871份,其中782份来自图书馆员,74份来自出版商,15份来自发现服务提供商[2]。根据此次调研结果,工作组针对当时存在的迫切性问题,决定在第一阶段制定以下5个方面的推荐实施规范:①数据格式、数据交换标准、传输方式以及持续更新;②向图书馆反馈受限内容的开放方式与程度;③明确与单项内容或集合内容索引级别、内容可用性相关的描述符;④定义从发现服务到内容的公平链接;⑤发现服务使用统计数据的采集内容、服务对象和数据传播途径[3]。

通过工作组成员的努力,2013年10月完成了推荐实施规范的草拟,并于2014年6月25日正式发布"开放发现倡议:提高发现透明度"(编号:NISO RP–19–2014)。该文档界定了发现服务相关术语,回顾了文献发现与传递服务模式的演进历程,分别针对内容提供商和发现服务提供商提出了最佳实施推荐规范,并列

chinaXiv:201803.00364v1

胡小丽，刘铁威，熊丽，等.提高开放发现透明度：美国 NISO ODI 述评[J].图书情报工作,2018,62(6):51－60.

出了下一步要开展的重点工作。

NISO ODI 正式发布之后,2014 年 10 月 27 日,ODI 决定成立常务委员会(Standing Committee)取代最初的工作组。常务委员依然由来自图书馆、出版社、发现服务提供商三方面的代表组成,哈佛大学图书馆的劳拉·莫尔斯(L. Morse)和 Sage 公司的莱蒂·康拉德(L. Conrad)出任联合主席。NISO ODI 也进入到新阶段。ODI 常务委员会的主要作用是促进 ODI 推荐实施规范的推广应用,并为内容提供商和发现服务商提供应用指导与支持,为所有发现平台的利益相关者建立一个持续论坛等。

近两年,ODI 常务委员会的主要工作集中在:①敦促内容提供商和发现服务提供商进行 ODI 推荐规范遵循确认并鼓励其公开发布确认清单。②加强与文摘、索引数据库提供商之间的沟通研讨。2017 年 1 月开始对此问题进行全面的问卷调研。③发布相关资料为图书馆提供 ODI 推荐实施规范应用指导:一方面提供模板发动图书馆向其内容提供商、发现服务提供商索求《ODI 项目支持确认清单》;另一方面将内容提供商已公开发布的确认清单提供给图书馆,指导其更好地在发现服务中利用这些确认项。④召开各种座谈、答疑沟通会,发布宣传小册子,出版相关研究论著,推广 ODI 项目。⑤与 COUNTER 项目组进一步沟通协作,在发现服务使用数据的收集与统计上进一步沟通研讨,为下一步工作的具体执行奠定基础[4]。

在前期工作的基础上,ODI 下一阶段的工作计划有:跨行业的协作研讨机制、应用程序接口(APIs)、受限内容在发现服务中的处理、需求驱动资源查找、无偏颇链接监测研究和与 COUNTER 标准的增强互动等。

3　ODI 项目的核心内容与特点

ODI 旨在增强新一代发现服务中内容资源覆盖的透明度,并向利益相关机构推荐实施一致性的内容交换机制。倡议由四部分组成:导言、发现服务与传递服务的演进、推荐实施规范和发展建议。其中,推荐实施规范是其核心内容。

推荐实施规范,是 ODI 针对当前发现服务实践中的问题提出的重要建议。为了提高基于中心索引仓储的发现服务的透明度,保持资源内容交换的一致性,该实施规范对内容提供商与发现服务提供商分别提出实践建议,规范了两者之间的协作流程。同时,也为图书馆提供了内容提供商发现服务参与度和发现系统功能的评估方法与参考标准。

3.1　倡议规范和共享元数据,增强索引中心仓储组建的透明度

对于基于中心索引仓储的发现系统而言,元数据的组建尤为重要。元数据可以简单地分为“薄”“厚”两种元数据。“薄”元数据(thin metadata)包含的字段非常少,例如来自于 CrossRef 的元数据。由于 CrossRef 是为 DOI 和 OpenURL 转换提供中立的解析和转换的公益组织,其包含的元数据只为生成 OpenURL 所用,只包含题名、作者、合作者、来源、标识等基本信息。典型的“厚”元数据(thick metadata),除了“薄”元数据所含的基本信息外,还包含了主题、摘要等信息[5]。如果说元数据的规模影响着发现服务对资源的覆盖广度,那么元数据的“薄”“厚”则影响着对资源的揭示度、检索结果的相关性排序、去重、分面揭示及进一步的数据挖掘,影响的是对资源的覆盖深度。

鉴于元数据的重要性和当前发现服务生态系统中相关的实际问题,基于索引的发现服务发展趋势要求内容提供商、发现服务提供商和图书馆合作以确保最广泛的资源能够通过发现平台得以充分揭示。内容提供商应该向发现服务提供商提供核心元数据以及隐含的完整的全文或原始内容,以便发现服务提供商建立索引,满足授权用户的需求。

由于数据流动透明度不足,就需要定义各数据点,明确什么样的元数据或者数据元素在何种条件下可被哪类用户获取。为此,ODI 的推荐实施规范明确了内容提供商传递给发现服务提供商的元数据内容以及授权用户可以访问的数据。

表 1 和表 2 中所列出的元数据元素所依据的是 KBART(Knowledge Base and Related Tools)元数据编码模式。该模式主要适用于连续出版物与专著题名,用以改进 A-Z 列表,链接解析和开放链接。在 ODI 中已经扩展至主题、文摘/描述,内容类型和内容格式等元素。这一延展旨在为各类资源内容提供最低揭示指标。

表 1 列出了 ODI 建议的由内容提供商向发现服务提供商所提供的最简元数据字段。

除了上述基本核心元数据元素外,ODI 推荐实施规范还提出了增强型元数据元素(见表 2)。在索引中纳入增强型元素并用于相关度排序会极大改善用户的文献发现体验,也有助于习惯使用可控词表的图书馆员或研究人员的检索。

chinaXiv:201803.00364v1

表 1 内容提供商所需提供的核心元数据元素[3]

字段名称	释义
题名	内容项主标题
著者	内容项的著者 每个作者应该按姓、名的顺序列出
出版商	内容项的出版商名称
卷	如适用,指资源卷数
期	如适用,指资源期数
页码	如适用,指资源页码
日期/日期范围	出版日期 对于连续出版物,应覆盖出版物的日期范围
内容项标识符	内容资源印本或在线版本的一个或多个标准标识(如 ISSN、OCLC 号、ISBN、DOI 等) 标识符前应当附加标识类型标签
集合题名	用于描述包含单项的出版物或连续出版物,如收录论文的期刊题名、包含单曲的 CD 专辑名称等
集合题名标识符	上述定义的集合题名的标准标识符(如 ISSN、OCLC 号、DOI 等) 该标识前应当附加标识类型标签
内容地址链接	开放链接或者直接指向某内容项全文的链接
开放获取标注	遵照 NISO 开放获取元数据和指示符小组的建议,如果某内容项为开放获取,应当用“free to read”标明,否则保留空白
全文标识	用“是/否”描述内容提供商是否允许发现服务提供商为构建索引的目的访问全文内容(或者非印刷载体的标准或高清版本) 预计未来当发现服务商描述某一题名或集合的索引覆盖范围时会向图书馆披露这一信息
内容类型	用于识别所描述的内容是何种类型,如文本、录像、录音等 建议使用 MARC21 头标中记录类型(06 字符位)可控词表的文本表述来描述本字段内容
内容格式	用于描述内容的自然属性,如专著、连续出版物、组成部分、集合等 建议使用 MARC21 头标中书目级别(07 字符位)可控词表的文本表述来描述本字段内容

表 2 内容提供商提供的增强型元数据元素[3]

字段名称	释义
索引数据	用来描述内容的一个或多个关键词,取自可控词表或任意词条
全文/脚本	对于文本内容而言,是整个文献。对于音频和视频内容来说,就是该资源中有声音内容的完整脚本
摘要/描述	对某内容的文本总结或者(非文本内容)对内容实体的描述

除了上述向发现服务提供商提供元数据信息外,ODI 还建议内容提供商向图书馆和订阅者披露其参与发现服务的程度。针对每个市场产品(期刊集合、文摘或索引数据库等),内容提供商需要披露其所提供内容的覆盖范围和内容深度,详情见表 3。

为了进一步提高发现服务的透明度,除了上述对内容提供商提出推荐建议外,同时要求发现服务提供商向图书馆公开其元数据信息,见表 4。

通过上述对内容提供商与发现服务提供商所提供的元数据内容的规范与增强透明度措施的建议,在一定程度上保障了元数据的规模、质量,提升了发现服务的效益。

3.2 倡议公平/无偏颇链接

在发现服务的测评与建设中,发现服务链接的公平性是一个非常重要的问题。链接的公平性关系到发现服务的检索结果是基于商业关系,还是基于最终用户需求或者图书馆的意愿。此问题直接关乎着发现服务的价值,牵连着内容提供商、发现服务提供商、图书馆及最终用户的直接利益。为此,ODI 提倡公平链接并得到相关行业的一致认可。ODI 建议发现服务推行以下协议来保障链接的公平性:

表 3 内容提供商提供给图书馆的元数据概览[3]

产品包	公开信息			
	范围	向发现服务提供商提供的用于索引的内容	内容提供商类型	内容接收方
期刊集合 电子书集合 图像集合 音频集合 视频集合 数据库 数据集	期刊:一个或者多个日期区间 其他产品:全部数据或者部分数据(描述范围)	适用于所提供内容的所有选项:元数据、索引数据、摘要/描述、全文	内容提供商提供市场产品时的身份,选项包括: (1)出版商、集成商; (2)文摘和索引服务提供商、开放获取知识库/机构知识库; (3)全文内容一般来自出版商、集成商、开放获取知识库或机构知识库。 文摘和索引记录一般来自文摘和索引服务提供商	可以使用内容的发现服务商,选项包括: (1)向所有发现服务提供商开放,即内容可供任何感兴趣的开放服务提供商收割; (2)按照协议提供给开放发现服务提供商(提供开放发现服务商列表)

(1)发现服务不应该基于业务关系,在生成检索结果、进行相关性排序或决定链接顺序所采用的方法上对内容提供商或产品(尤其是其自有产品)予以差别对待。

(2)在多个内容平台中存在相同内容(如一个主要出版商和一个辅助数据库,或多个辅助数据库)的情况下,发现服务提供商应该提供机制,使图书馆能够干预向用户呈现哪些平台以链接到目标内容以及按照何种次序呈现。

chinaXiv:202308.00364v1

表4　发现服务提供商提供给图书馆的元数据概览[3]

字段名称	释义
出版物名称	内容题名
日期/日期范围	出版日期;连续出版物使用日期范围
标识符	出版物印本或在线版本的一个或多个标准标识符(如 ISSN、OCLC、ISBN、DOI 等)。标识符前应当有标识类别标签
内容格式	用于表示所描述的内容的性质,如专著、连续出版物、组件、合集等 建议使用 MARC21 头标中书目级别(07 字符位)可控词表的文本表述来描述本字段内容
内容类型	用于识别所描述的内容是何种类型,如文本、录像、录音等 建议使用 MARC21 头标中记录类型(06 字符位)可控词表的文本表述来描述本字段内容
被索引的内容	适用于被索引内容的选项:元数据、索引数据、摘要/描述、全文

(3)发现服务提供商应该每年向其所有客户(主要为图书馆)发布报告或者在其网站公开发表声明,说明其与内容提供商的业务联系,包括那些存在直接或间接从属关系的内容提供商以及那些已经针对元数据、直接链接以及其他特殊安排与之达成有偿协议或其他协议的内容提供商。

(4)发现服务提供商应该提供一个声明,确认在涉及理念和政策、信息来源以及任何其他相关因素时,其用于生成结果集、相关度排序以及链接次序的算法都具有中立性。

(5)发现服务提供商应该将与某一检索结果相关的、涉及链接呈现方式的相关部分设置为图书馆可选择的参数配置选项(包括显示的链接数目、检索结果的链接排序以及图书馆如何标识"获取全文"的链接等)。

(6)当文摘和索引提供商能够提供其链接时,发现服务提供商应该提供从发现界面到发现服务索引所使用的文摘和索引界面的无缝链接(不一定是全文链接)。

(7)当发现服务的内容资源变化可能会影响到检索结果集、相关度排序或结果的链接顺序时,发现服务提供商应当向内容提供商和图书馆提供相关信息。

3.3　规范数据格式与传输标准,保障内容交换的一致性

数据格式与传输方式无疑是发现服务中重要的实操性基础工作。ODI 提倡数据交换一致性原则,规范数据格式与数据传输标准,降低发现服务生态系统中内容流动的复杂度,降低那些可能会阻碍内容提供商广泛参与的技术门槛。ODI 提议内容提供商向发现服务提供商传输数据时应遵循以下数据传输机制与协议:

(1)在关于内容提供商应该如何向发现服务提供商传输数据以便发现服务最有效地构建索引方面,发现服务提供商应该清楚地描述其性能、局限和偏好。

(2)当格式、模式或传输机制将会影响到发现服务的特色或性能时,发现服务提供商应当与内容提供商进行沟通。

(3)元数据编码模式方面,建议尽可能地使用稳健的编码模式来描述内容提供商的元数据。内容提供商和发现服务提供商应当就适当、可支持的编码模式达成一致,并最好向图书馆披露。在图书馆和出版业中广泛使用的主要的元数据编码模式有 MARC、MODS、METS、VRA、DC、KBART、EAD 和 ONIX;推荐的文本格式有 CSV、TXT、XML 等;推荐的传输方式有 oai-pmh、resourcesync、api 和 ftp。

ODI 这一提议有助于解决现有发现系统与商用数字资源数据的规范性与标准化问题,特别是对于国内标准化建设缺失的一些中文数据库和自建数据库有着重要的借鉴意义。通过一个推荐建议的实施,有利于提高其资源在发现服务中的被整合度与被发现度。

3.4　重视数据统计,推荐最低统计指标

使用统计分析,无论对于内容提供商来说还是图书馆来说,其重要性都是不言而喻的。发现服务提供商应该向内容提供商和图书馆分别提供一组结构化的指标以说明其资源在发现服务的使用情况,这一提倡有助于鼓励内容提供商加大发现服务的参与度,同时,也有利于图书馆监测发现服务中的使用问题,及时提出配置与功能调整需求,提高发现服务效益。

ODI 建议发现服务提供商向资源内容提供商提供以下指标性数据:

(1)检索总数:检索是用户从搜索框发起的一个查询请求。需要注意的是,依据 COUNTER 标准,针对原始查询请求采取的筛选、排序或者换页操作不应该计数,这些改变应该被记录为操作,而不是检索。此外,任何发现服务创建的自动搜索,都不应该计入检索总数。

(2)结果点击数:发现服务提供商应向内容提供商提供最终用户对检索结果中其所属资源的点击次数。需要注意的是,当有多个内容来源为发现服务提供文章的元数据时,每个内容提供商都应记入一个结果点击增量。此外还要注意,不论一个用户点击同一条记录多少次,一个结果只能计数一次。

chinaXiv:202308.00361v1

ChinaXiv合作期刊

(3)点入总量:该指标报告某一内容提供商的全文内容被最终用户访问的频率。发现服务只能记录对内容的请求数量,但不能监测有多少内容被成功获取。点入总量指标还可以提供有关用户请求来源的信息,即通过 OpenURL 访问还是通过发现服务提供的直接链接访问。

对于图书馆而言,关于使用统计的指标,ODI 建议发现服务提供商至少需要提供以下数据:

(1)每月检索总数:该指标记录了发现服务中用户检索总次数,可以跟踪每月的使用情况趋势。此外,发现服务提供商还应该尽可能地针对文摘和索引数据库以及类似资源提供每个内容集合的检索次数,因为这些资源通常只有认证用户才能访问。

(2)每月总访问数:该指标记录了每月访问人数,向图书馆提供该馆用户的月度使用趋势。

(3)每月点入总数:该指标汇报的是用户点击某一链接请求获取全文内容的总次数。这一指标显示了发现服务提供的有价值的结果频率。

(4)最近前 500 个检索词条:该指标指特定时间段内前 500 个使用频率最高的检索词条。此数据有助于图书馆对读者检索类型与检索主题信息的收集与挖掘。

(5)近期跳转到发现服务的前 100 个推荐链接:该指标指的是用户链接到发现服务最常用的 100 个链接。这一指标用来向图书馆显示用户接入发现服务的渠道。

ODI 关于使用统计指标的建议虽然略显粗犷,每一个建议指标项还需进一步细化,但是这些指标的提出在一定程度上规范了发现服务提供商的使用统计功能与反馈机制,提高了内容提供商对发现服务的参与度,同时也有利于图书馆掌握其馆藏资源利用情况,并通过相关统计数据的分析、挖掘获取读者使用行为数据,提升其智能检索与文献推荐服务水平,从而实现发现服务生态系统的良性运作。

4 ODI 推荐实施规范应用实践

为了推进 ODI 推荐实施规范在行业里的实际应用,ODI 分别为内容提供商和发现服务提供商制定了《ODI 项目支持确认清单》模板,并鼓励他们公开发布《确认清单》,以增强发现服务的透明性,促进发现服务的良性发展。

4.1 内容提供商与 ODI 推荐实施规范

ODI 为内容提供商所制订的支持确认清单主要有元数据共享、提供发现服务参与度说明、无保密条款声明和规范数据格式与传输标准等 4 个方面 6 项内容,每一项都有符合、部分符合和不符合 3 个指标。如表 5 所示:

表 5 内容提供商 ODI 项目支持确认清单模板[3]

符合/不符合/部分符合	推荐规范	参考章节	备注
	内容提供商向发现服务提供商提供核心元数据、可能的全文/原创内容等完整数据内容	3.2.1.1 (1)	
	针对提交给索引的每一单项,内容提供商提供给发现服务提供商核心元数据组成元素(见 3.2.1.2)	3.2.1.1 (2)	
	内容提供商尽可能多地提供其内容项以及附加描述内容	3.2.1.1 (3)	
	内容提供商应图书馆的请求,提供参与发现服务的说明,包括披露其内容覆盖度与内容深度	3.2.2	
	内容提供商与发现服务提供商之间的协议不包含保密条款	3.2.3	
	内容提供商向发现服务提供商进行数据传输时尽量使用已有标准,并且使用 3.3.3 中所列的元数据编码方案	3.2.4	

截至 2017 年 9 月,公开发布完整的 ODI 项目支持确认清单的内容提供商有 CREDO、EBSCO、Emerald、IEEE、Gale 和 SAGE 6 家。

在元数据共享方面,6 家内容提供商都表示针对提交的每一单项索引会提供其核心元数据,并会尽可能多地提供其他内容项与附加描述;但是在尽可能完整提交其所有内容的核心元数据和全文及原创数据时,6 家提供商支持度略有不同。CREDO[6]、Emerald[7] 和 IEEE[8] 3 家表示完全支持,EBSCO、Gale 和 SAGE 3 家部分支持。EBSCO 既是内容提供商亦是发现服务提供商,目前,通过其数据共享与协作政策,向主要发现服务商提供了 262 个全文数据库及其他资源,但是其文摘和索引数据库不包含在政策之内[9];SAGE 统计目前暂不支持发现服务索引[10],Gale 针对不同的发现服务提供商所提供的可索引内容和全文链接方式有所不同[11]。

在提供内容覆盖广度与深度说明方面,上述 6 家提供商都提供了详细的资源说明清单。如图 1,IEEE Xplore 向图书馆的说明中提供了可被发现服务索引到的资源及其覆盖起止年限、所提供索引的类型、内容提供商类型、提供给哪些发现服务提供商等信息[8]。

chinaXiv:202308.00364v1

Market Product	Coverage Start	Coverage End	Content Provided to DSP for Indexing	Content Provider Type	Content Provided to
IEEE Journals	1884	Present	Full Text or Metadata	Publisher	Available to all Discovery Service Providers
IEEE Conference Proceedings	1951	Present	Full Text or Metadata	Publisher	Available to all Discovery Service Providers
IEEE Standards	1949	Present	Full Text or Metadata	Publisher	Available to all Discovery Service Providers
IEEE Wiley eBooks	1974	Present	Full Text or Metadata	Publisher / Aggregator	Available to all Discovery Service Providers
(IEEE) MIT Press eBooks	1943	Present	Full Text or Metadata	Aggregator	Available to all Discovery Service Providers
(IEEE) Morgan & Claypool eBooks	2006	Present	Full Text or Metadata	Aggregator	Available to all Discovery Service Providers
(IEEE) SMPTE Digital Library	1916	Present	Full Text or Metadata	Aggregator	Available to all Discovery Service Providers

图 1　IEEE 提供的组建发现服务索引资源清单

chinaXiv:202308036v1

目前,上述 6 家内容提供商均向主要的图书馆发现服务系统提供了内容,如 ProQuest （Summon）、EBSCO Discovery Service （EDS）、Ex Libris （Primo）、OCLC WorldCat Discovery、TDNet,以便其建立索引。此外,它们还不同程度地推行了许多有助于发现服务的举措。例如 Gale 支持 Open URL 链接,允许从外部信息源链接到 Gale 数据库的全文。这样,来自图书馆发现服务、文摘和索引数据库以及其他使用 Open URL 链接的研究人员都能更好地访问 Gale 的资源[11]。SAGE 也在 4 个关键领域解决其产品的可发现性问题:开放网络发现、机构发现、专业数据库发现和跨 SAGE 发现,从而确保最终用户无论从哪里开始检索,使用何种设备都能够顺畅有效地获取 SAGE 内容。SAGE 还与时俱进地完善了其 MARC 记录政策:免费提供电子书和参考产品(包括视频、数据集、案例研究)的 MARC 记录,客户可以从 SAGE 指定的若干平台上下载。符合 MARC21 格式的记录包括 RDA 元素,从而推动了 SAGE 资源在图书馆的发现与使用[10]。

在数据格式与传输方面,各家内容提供商都表示完全遵循 ODI 最佳推荐指导,根据 KBART 并且依照其他标准格式(XLS、XLSX、HTML、CSV 和 XML)提供多数产品的题名清单。

4.2　发现服务提供商与 ODI 推荐实施规范

ODI 为发现服务提供商所提供的支持确认项主要包括内容列表、链接说明和向内容提供商与图书馆提供使用统计报告 4 个方面,共 16 项,每个确认项有符合、不符合和部分符合 3 个指标项,如表 6 所示:

表 6　发现服务提供商 ODI 项目支持确认清单模板[3]

符合/不符合/部分符合	推荐规范	参考章节	备注
	为图书馆提供内容列表	3.3.1	
	链接	3.3.2	
	发现服务在链接内容提供商产品时不存在偏颇	3.3.2 (1)	
	提供参数调整机制,允许图书馆通过该机制决定哪些平台为目标链接以及排列次序与优先权	3.3.2 (2)	
	在内容索引与结果推送方面,发现服务确认无偏颇。有关此方面的声明应每年发布	3.3.2 (3)	
	发现服务在生成结果集、进行相关度排序以及决定链接次序上所用算法对于内容提供商没有优先特惠照顾	3.3.2 (4)	
	与既定结果关联的链接可被图书馆进行配置	3.3.2 (5)	
	提供从发现界面到发现服务索引所使用的文摘索引界面之间的无缝链接(并非一定是全文链接)	3.3.2 (6)	
	向内容提供商提供使用情况报告	3.3.4.1	
	提供检索总数报告	3.3.4.1.1 (1)	
	提供结果点击数报告	3.3.4.1.1 (2)	
	提供点入总量报告	3.3.4.1.1 (3)	
	提供发现服务接入来源网址报告	3.3.4.1.2	
	向图书馆提供使用情况报告	3.3.4.2	
	提供每月检索总数报告	3.3.4.2 (1)	
	提供每月访问总数报告	3.3.4.2 (2)	
	提供每月点入总量报告	3.3.4.2 (3)	
	提供近期排名前 500 的检索词报告	3.3.4.2 (4)	
	提供近期跳转到发现服务的 100 个最常用推荐链接报告	3.3.4.2 (5)	

ChinaXiv合作期刊

在 ODI 推荐实施规范正式发布之后,EBSCO 作为权威的学术资源内容提供商和主要的发现服务提供商于 2014 年 6 月 26 日公开发表声明宣布支持 ODI 项目,倡议协作、元数据共享和推动发现服务透明度。之后 Ex Libris(Primo)于 2015 年 6 月 24 日、ProQuest (Summon)于 2015 年 6 月 26 日发布声明支持 ODI 项目,并相继公开了其 ODI 项目支持确认单。

根据 EBSCO、Ex Libris 和 ProQuest 3 家公司提供的确认清单,他们目前支持 ODI 项目的情况如下:

在向图书馆提供内容列表方面,EBSCO 完全支持,Ex Libris 和 ProQuest 支持提供部分内容清单。例如,Ex Libris 目前提供集合(collection)层面的列表,未提供题名层面的列表。

在发现服务链接方面,3 家公司都宣称支持公平链接,不会偏向任何内容提供商或内容产品。在允许图书馆干预链接目标与排列次序上,EBSCO 和 Ex Libris 表示完全符合 ODI 推荐规范。Ex Libris 声称致力于提供顺畅有效的发现体验,向用户提供最相关的内容与检索结果。并承诺坚持中立原则,与内容提供商没有排他性协议,对任何平台与提供商均无歧视,其搜索结果排序、内容链接、Primo Central 收录的资料以及发现开放获取的内容,都秉持内容中立的原则[12]。ProQuest 声称部分符合规范,认为图书馆可以设定偏好,但是 Summon 系统在不偏向任何内容提供商的前提下,为避免某些链接无法访问,会兼顾考虑到“最佳可访问链接”[13]。

在向内容提供商和图书馆提供使用报告方面,3 家系统都可以提供检索总量、结果点击量、点入总量、接入链接和检索词等相关的报告,但是报告的详略情况不一。例如:在每月的点入总量方面,ProQuest 的发现系统目前可以按内容提供商提供点入数据报告,但是还暂时不能提供特殊集合内容的点入次数报告。EBSCOadmin 可以提供多个阶段性报告,但是其所提供的报告仍然重复计算来自同一 IP 地址以及来自特定个人访问者的多次访问。此外,EBSCO 目前无法完全提供显示每月用户数量的专项报告。但是,EBSCO 的发现服务可以嵌入到 Google Analytics 跟踪分析中,可以用其中使用会话/访问者数量作为其用户数的替代性报告。在排名前 500 的检索词反馈方面,EBSCO 表示部分支持。近期排名前 500 的检索词目前还没有作为标准的 EBSCO 发现服务报告内容提供给所有客户,但是,EBSCO 发现服务可以嵌入 Google Analytics,跟踪分析检索词,可以为用户提供高频检索词报告。在清单的更新确认方面,EBSCO 表示会根据进展及时更新 ODI 项目支持确认清单中的相关表述[14]。

4.3 图书馆与 ODI 推荐实施规范

基于中心索引仓储的新一代发现服务,以上一代联邦检索所无法比拟的便捷、快速、高效的检索体验,很快成为学术性图书馆发现服务的主体模式,这一点由图书馆对新一代发现系统的引进即可管窥一般。

早在 2012 年,ODI 项目启动之初的调研数据就显示了发现服务在图书馆文献服务中的重要地位。根据 ODI 调研数据(见图 2),在 782 份来自图书馆的调研数据中,当被问及发现系统的使用情况时,其中 74% 的图书馆显示已在使用发现系统,17% 的图书馆显示会在未来的 1 -2 年内使用发现系统。根据 *Library Journal* 的统计数据,截至 2011 年底,就已经有 4 128 个[15]发现服务被应用。

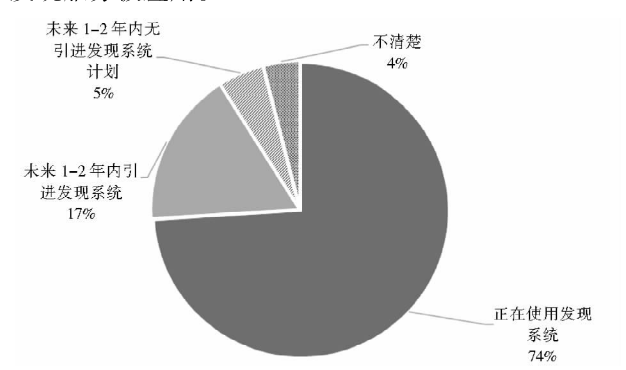


图 2 发现系统使用情况调查

国内图书馆在 2009 年后也陆续引进发现系统推行服务。以引入国际三大商业发现系统为例,截至 2017 年 9 月,国内引入 Primo 的图书馆有 52 家[16],引入 Summon 的图书馆有 54 家[16],引入 EDS/Find + (EDS 进入中国后与江苏宝和数据股份公司技术合作衍生出 Find +)的图书馆有 78 家[17],基本涵盖了“985”“211”等重要高校图书馆与研究院图书馆。

图书馆在发现服务中的角色与权益是不言而喻的,ODI 项目委员会发布指导说明倡议图书馆选择发现服务,敦促内容提供商与发现服务提供商遵循推荐实施规范,提高发现服务透明度。在国外,东亚图书馆、哈佛大学图书馆等都成立了专门小组召开会议讨论、号召发现服务 ODI 支持事宜。在 2017 年 5 月 Ex Libris 发现服务的北美用户年会上,ODI 常务委员会联合主席、哈佛大学图书馆系统部主任 L. Morse 号召用户协会联合敦促增强发现服务透明度,支持 ODI 实施规范,提出图书馆应该:①倡议并积极参与开放获取;②与出版商、经销商一起发起专业会议研讨协作利益、

chinaXiv:202303.00364v1

发现系统现状和学术链需求等;③跟踪并参与标准建设,特别是在 ODI 实施推广方面;④将经费用在值得投入的地方。并对图书馆近期可开展的工作提出了具体的指导:①分析文摘与索引数据库在发现系统中缺失的实际问题;②以可视化图谱的方式展示出因文摘与索引资源缺失所造成的影响,通过与文摘和索引提供商的协作洽谈,最终实现 Summon 和 Primo 等对此类资源的发现服务;③向参考馆员和其他职员推荐 ODI 实施规范和相关工具,最大化地优化发现系统[18]。

目前,ODI 项目还未在国内图情界得到推广,但是在发现服务建设中,国内图书馆在发现系统选型、资源配置等相关工作中所依据的原则却与 ODI 项目目标殊途同归。馆藏资源覆盖度(特别是对中文资源的覆盖度)、元数据的来源规范及"厚""薄"度(元数据揭示的层级)、检索机制、数据转换与传输、使用统计、全文获取、与其他系统平台数据对接、检索结果相关性排序[19-20]等方面都是国内图书馆所考虑与测评的重点,也是发现系统建设与服务推行中问题与困扰较多之处。ODI 项目的引进与推广,将会为国内图书馆相关开放发现的呼吁以及中文数据库商的数据标准化建设提供原则依据与方法指导,为国内发现服务的推行与评估提供实践指导。

5 总结与瞻望

综上所述,基于中心索引仓储的发现服务模式极大地提升了图书馆的资源整合与文献发现服务水平,成为国内外学术研究型图书馆提供馆藏文献资源检索与发现的主体服务模式。其所涉及的元数据规范与开放共享、公平链接、数据格式与传输标准、使用统计等方面的问题都是国内外发现服务所面临的共性问题。ODI 从解决实际困难出发,通过增强新一代发现服务重要环节的透明度,协调内容提供商、发现系统提供商和图书馆三方权益,为国内外发现服务的实践提供了解决方法与处理依据,推动了新一代发现服务的进程。

当然,在发现服务这个复杂的生态系统中,还存有许多实际困难需要解决。例如跨行业利益相关者之间的协作机制、应用程序接口(APIs)、文摘与索引资源在发现服务中的使用、使用统计数据的细化、检索结果相关度排序等。此外,在开放发现服务的发展与延伸上也有许多前瞻性问题需要进一步探讨。例如,发现服务系统数据挖掘及其与其他应用程序的数据对接、交换、开放使用,基于发现服务统计的个性化智能推荐服务,基于语义、概念推理、人工智能的知识发现系统(基于人工智能的 Yewno 知识发现系统目前与斯坦福大学图书馆合作测试运行)构建等问题。这些问题有的已经纳入到 ODI 下一步的工作计划中,有的已引起业界和相关领域的关注。笔者相信,随着对发现服务及其拓展应用的不断聚焦,随着新技术、新理念在发现服务中的创新应用,与时俱进的 ODI 项目将会跨行业深度践行,从而推动发现服务的持续性发展,促进数字资源建设与数字化服务的规范化与标准化进程。

参考文献:

[1] BREEDING M. Automation marketplace 2013: the rush to innovate [EB/OL]. [2017-06-10]. http://www.thedigitalshift.com/2013/04/ils/automation-marketplace-2013-the-rush-to-innovate.

[2] NISO ODI working group. ODI survey report: reflections and perspectives on discovery services [EB/OL]. [2017-08-19]. http://www.niso.org/apps/group_public/download.php/9977/NISO ODI Survey Report Final.pdf.

[3] Open discovery initiative working group. Open discovery initiative: promoting transparency in discovery [EB/OL]. [2017-06-30]. http://www.niso.org/publications/rp/rp-19-2014/.

[4] ODI standing committee. ODI update [EB/OL]. [2017-08-10]. http://www.niso.org/workrooms/odi/updates/.

[5] 窦天芳,姜爱蓉. 资源发现系统功能分析及应用前景[J]. 图书情报工作, 2012, 56(7):38-43.

[6] Credo. Discovery services and metasearch vendors [EB/OL]. [2017-08-10]. http://corp.credoreference.com/partners/discovery.html.

[7] Emerald publishing limited. ODI checklist [EB/OL]. [2017-08-10]. http://www.emeraldgrouppublishing.com/support/files/odi_checklist.docx.

[8] IEEE xplore. Open discovery initiative conformance [EB/OL]. [2017-08-10]. http://ieeexplore.ieee.org/Xplorehelp/#/tools/toolsForAdministratorsAndLibrarians/discoveryServices/openDiscoveryInitiativeConformance.

[9] EBSCO. EBSCO ODI conformance checklist [EB/OL]. [2017-08-10]. https://www.ebscohost.com/odi-conformance-checklists.

[10] Sage publication. Indexing & discoverability [EB/OL]. [2017-08-10]. https://uk.sagepub.com/en-gb/asi/indexing-discoverability#OpenDiscovery.

[11] Gale. Gale discovery resources [EB/OL]. [2017-08-10]. http://www.gale.com/discovery.

[12] Ex Libris. Open discovery initiative conformance [EB/OL]. [2017-08-10]. http://www.exlibrisgroup.com/default.asp?catid=%7B8B84715C-234A-44A9-B3C1-DCDD2E18D049%7D.

[13] ProQuest. ODI conformance checklist: discovery service provider [EB/OL]. [2017-08-10]. http://media2.proquest.com/documents/NISO+ODI+conformance+checklist+-+Discovery+Service+Provider2.pdf.

chinaXiv:202308.00364v1

ChinaXiv合作期刊

[14] EBSCO. ODI conformance checklist: discovery service provider [EB/OL]. [2017 -08 -10]. https://www.ebscohost.com/odi-conformance-checklists.
[15] The digital shift. Automation marketplace 2012: the complete survey data [EB/OL]. [2017 -08 -10]. http://www.thedigitalshift.com/2012/03/ils/automation-marketplace-2012-the-complete-survey-data/#2011-sales.
[16] 艾利贝斯软件科技发展(北京)有限公司. CCEU 用户协会 [EB/OL]. [2017 -06 -10]. http://www.exlibris.com.cn.
[17] 江苏宝和数据股份有限公司. 资源发现 EDS/Find + 用户协会 [EB/OL]. [2017 -08 -10]. http://www.smartlib.cn/do/list.php? fid =32.
[18] MORSE L, AARON A. Increasing depth and scope of content in discovery library advocacy discussion [EB/OL]. [2017 -08 -10]. http://www.niso.org/apps/group_public/download.php/18061/ELUNA2017-Increasing depth and scope of content in discovery.pptx.
[19] 朱前东. 资源发现系统评价体系构建研究[J]. 图书馆建设, 2014(4):41 -45.
[20] 孙宇,张磊,刘炜. 图书馆资源发现系统选型研究[J]. 图书馆杂志, 2013,32(12):63 -70.

作者贡献说明:

胡小丽:负责论文大纲拟定、资料与数据收集、论文的撰写与修改、审核;
刘铁威:负责论文相关资料与数据的收集、论文审核与校对;
熊丽:负责论文审核与校对;
薛燕:负责论文审核与校对。

Review on Open Discovery Initiative: Promoting Transparency in Discovery

Hu Xiaoli[1] Liu Tiewei[2] Xiong Li[1] Susan Xue[3]
[1] Capital Normal University Library, Beijing 100089
[2] Henry Madden Library, California State University, Fresno 93740 -8014
[3] C. V. Starr East Asian Library, University of California, Berkeley, Berkeley 94720 -6000

Abstract: [**Purpose/significance**] The purpose of studying the Open Discovery Initiative (ODI): Promoting Transparency in Discovery released by National Information Standards Organization (NISO) is to promote the in-depth understanding and extensive implementation of ODI in libraries in China, to enhance the development of current generation discovery service based on centrally index and improve the standardization of digital resource in China. [**Method/process**] The paper mainly introduces the initiation, development, key content and characteristics of ODI, specially highlights the recommended practices and discusses its implementation status. Further, the paper discusses the current implementation status and the future development of ODI. [**Result/conclusion**] DOI aims to promote transparency in discovery and facilitate progress through exploration of metadata elements, disclosure, fair linking, technical formats, and usage statistics, etc. More and more content providers and discovery service providers have gradually accepted and complied with ODI. That isbenefit to the improvement of discovery service and standardization of digital resource.

Keywords: open discovery initiative discovery service central index metadata

chinaXiv:202308.00364v1

《图书情报工作》投稿作者学术诚信声明

《图书情报工作》一直秉持发表优秀学术论文成果、促进业界学术交流的使命,并致力于净化学术出版环境,创建良好学术生态。2013 年牵头制订、发布并开始执行《图书馆学期刊关于恪守学术道德净化学术环境的联合声明》(简称《声明》)(见:http://www.lis.ac.cn/CN/column/item202.shtml),随后又牵头制订并发布《中国图书馆学情报学期刊抵制学术不端联合行动计划》(简称《联合行动计划》)(见: http://www.lis.ac.cn/CN/column/item247.shtml)。为贯彻和落实这一理念,本刊郑重声明,即日起,所有投稿作者须承诺:投稿本刊的论文,须遵守以上《声明》及《联合行动计划》,自觉坚守学术道德,坚决抵制学术不端。《图书情报工作》对一切涉嫌抄袭、剽窃等各种学术不端行为的论文实行零容忍,并采取相应的惩戒手段。

《图书情报工作》杂志社
2017 年 11 月 28 日